

MITOCHONDRIAL PROTEOMICS OF A SECONDARY GREEN ALGA

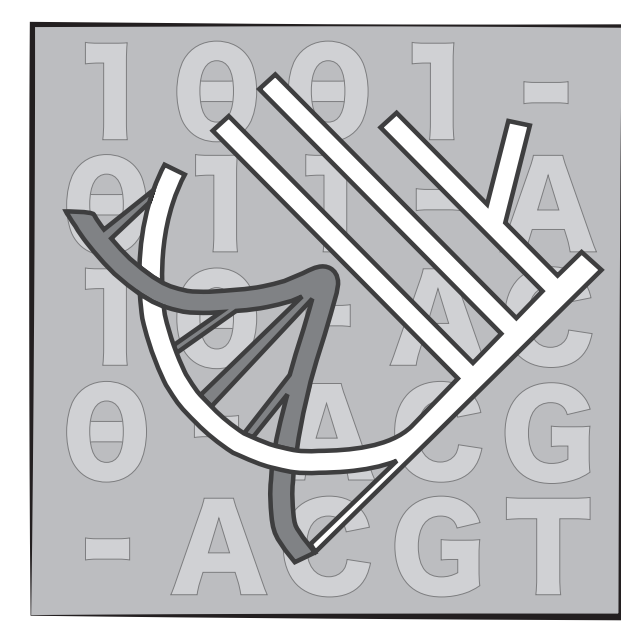


Emilie PEREZ^{1,2}, Hervé DEGAND³, Pierre MORSOMME³, Claire REMACLE²,
Pierre CARDOL², Denis BAURAIN¹

¹Phylogénomique des Eucaryotes, Institut de Botanique, Université de Liège, Belgium

²Génétique des Microorganismes, Institut de Botanique, Université de Liège, Belgium

³SST/ISV, Institut des Sciences de la Vie, Université Catholique de Louvain, Belgium



Euglena gracilis, a secondary green alga

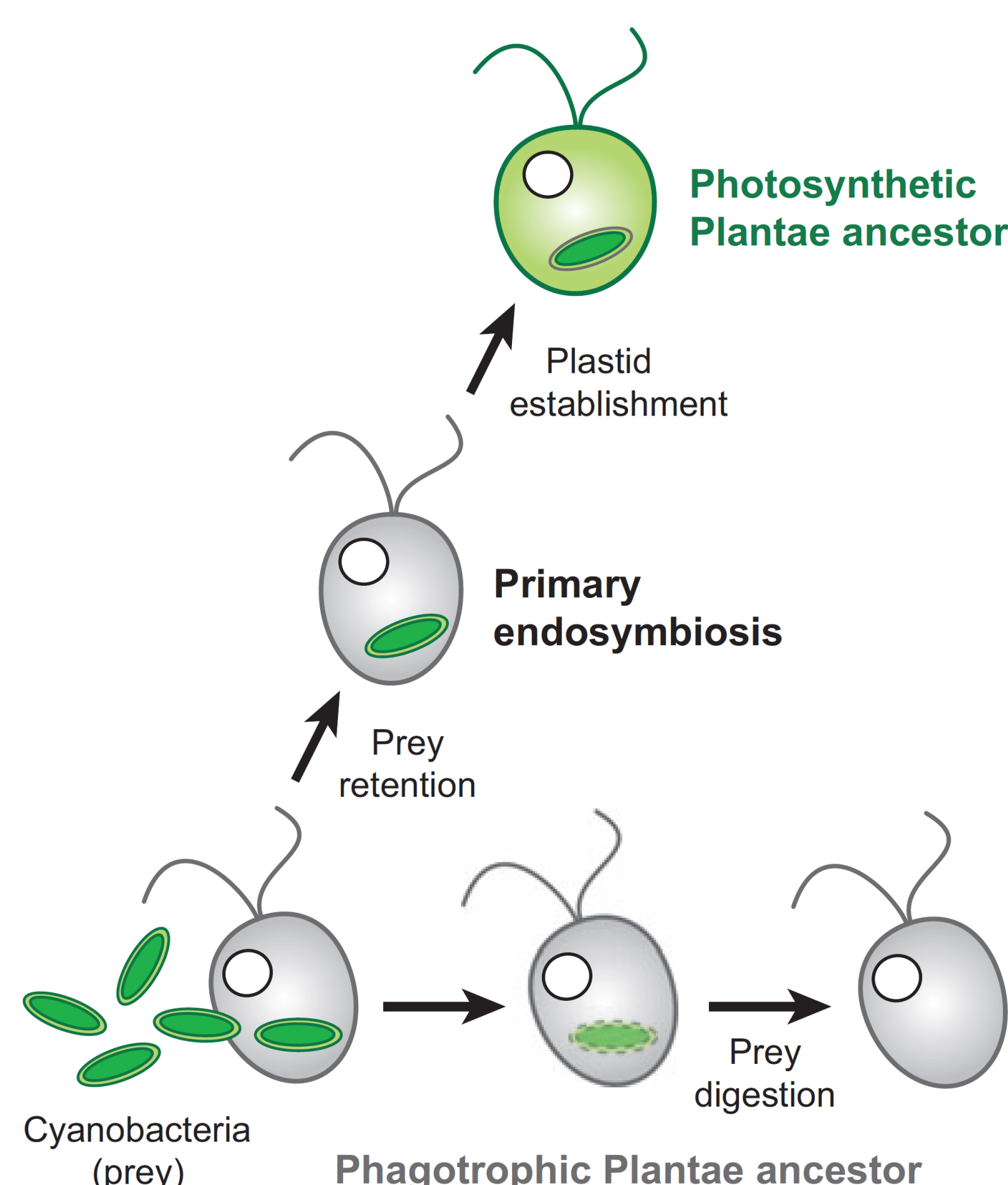


Figure 1: Primary endosymbiosis. Origin of photosynthetic cells.

DURING evolution, some eukaryotic lineages have acquired the ability to photosynthesize through endosymbiosis and conversion of the endosymbiont into a plastid. The endosymbiosis is termed primary when the endosymbiont is a cyanobacterium whereas it is termed secondary when it is a eukaryotic alga. *Euglena gracilis* derives from a secondary endosymbiosis between its phagotrophic ancestor and a green alga. It is thus a kind of repetition of the much older primary endosymbiosis.

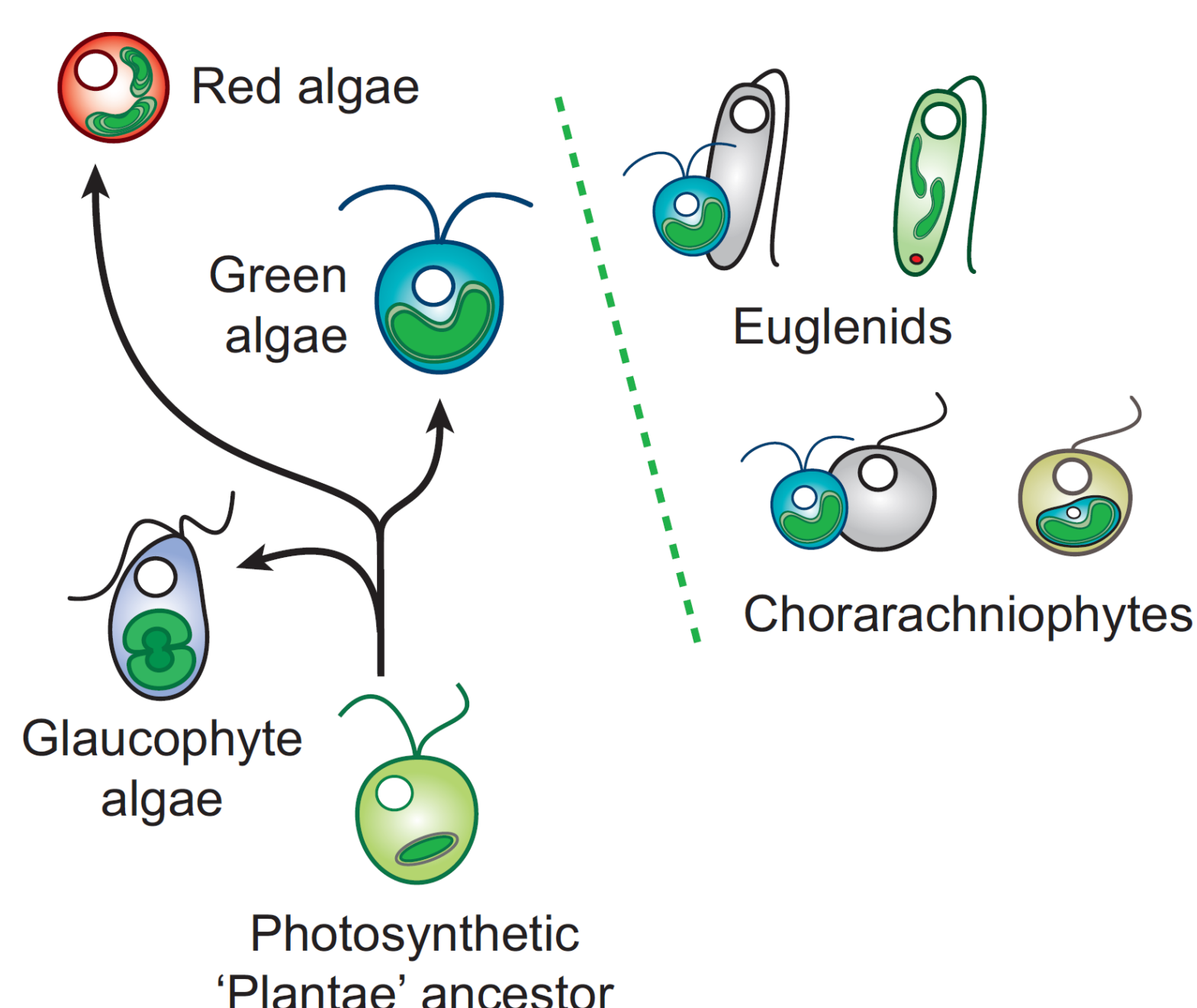


Figure 2: Secondary endosymbiosis. Origin of Euglenids.

CONSIDERED as a secondary green alga, *Euglena gracilis* belongs to the phylum of euglenids, which are part of the larger group Euglenozoa (supergroup Excavates). Also included in Euglenozoa are kinetoplastids, which encompass trypanosoma (the parasites causing sleeping sickness, among other diseases). Algae such as *Euglena gracilis* are termed "complex" because they are genetic chimeras. Due to their particular history, their genes may come from Euglenozoa but also from the green lineage (or possibly from other lineages such as red algae).

THE principal metabolic pathways, especially those ensuring cellular energetics, are mainly found in the plastid and the mitochondrion. Our general objective is to study the interactions established between these two organelles during secondary endosymbiosis and to determine the phylogenetic origin of the genes encoding proteins involved in these interactions. As a first step, we performed a high-throughput analysis of the mitochondrial proteome of *Euglena gracilis*.

Methods

THE mass spectrometry analysis of proteins in samples of mitochondria extracted from *Euglena gracilis* allowed us to identify several hundred protein fragments. First, the peptides obtained by MS/MS were compared to a translated database of publicly available *Euglena gracilis* ESTs. Then, the ESTs matching one or more peptides were annotated with BLAST against the NCBI non-redundant protein database. This strategy led to the identification of most protein fragments obtained by mass spectrometry.

Subcellular localization of identified protein fragments

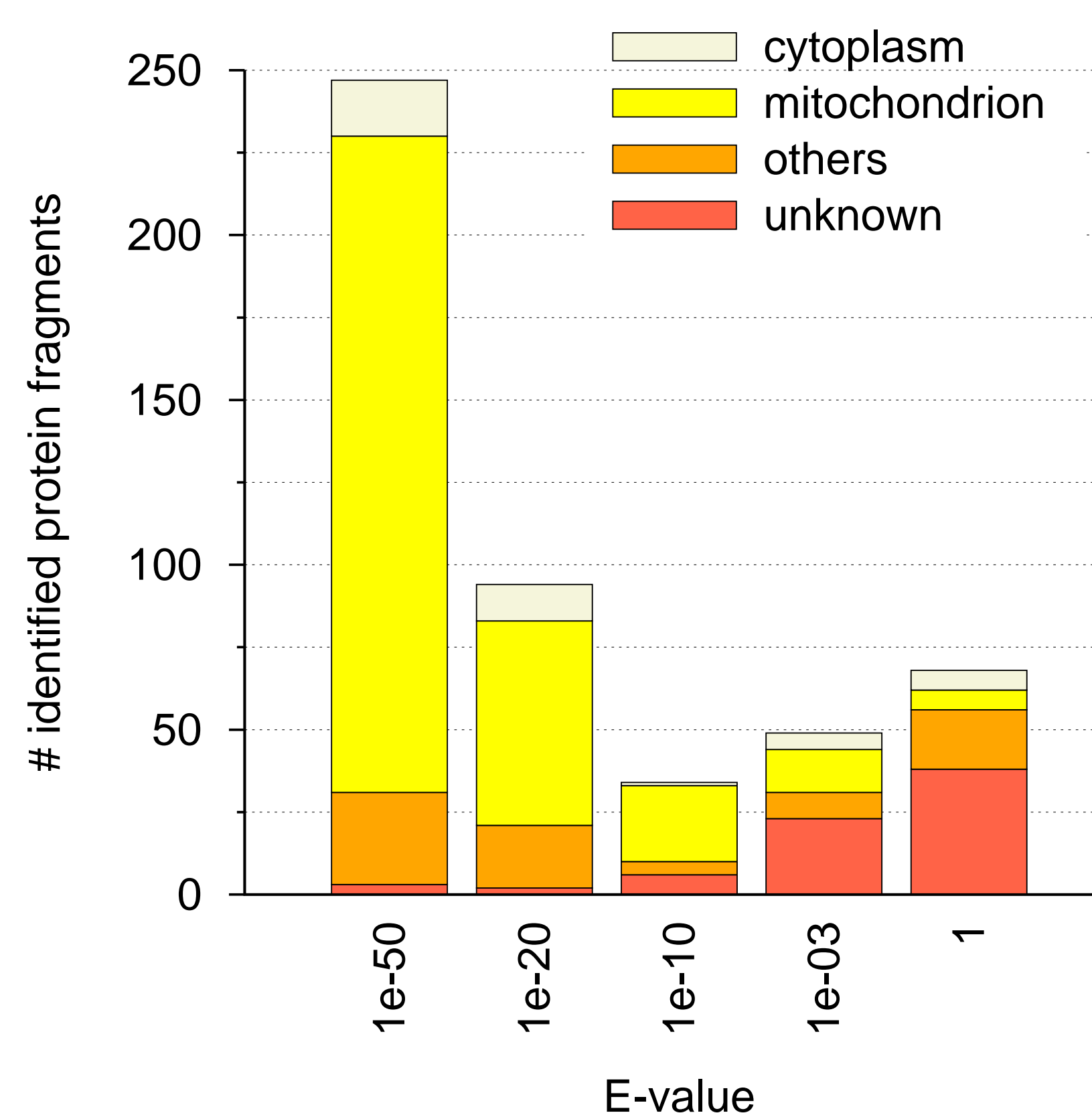


Figure 3: Subcellular localization of protein fragments broken by E-value. 1e-50: $E\text{-value} \leq 1e-50$, 1e-20: $1e-50 < E\text{-value} \leq 1e-20$, 1e-10: $1e-20 < E\text{-value} \leq 1e-10$, 1e-03: $1e-10 < E\text{-value} \leq 1e-03$ and 1: $1e-03 < E\text{-value} \leq 1$.

WE can evaluate the quality of our analysis by focusing on the subcellular localization of these protein fragments. Here, we see that for the best E-values ($< 1e-10$), the vast majority of protein fragments are localized in the mitochondrion. In contrast, for low ($> 1e-10$) or non-significant E-values ($> 1e-03$), the mitochondrial subcellular localization is much rarer, thus indicating that the identification is probably incorrect. Furthermore, we observe that there are very few identifications with non-significant E-values (about 70 of nearly 500 identified protein fragments).

Taxonomic affiliation of identified protein fragments

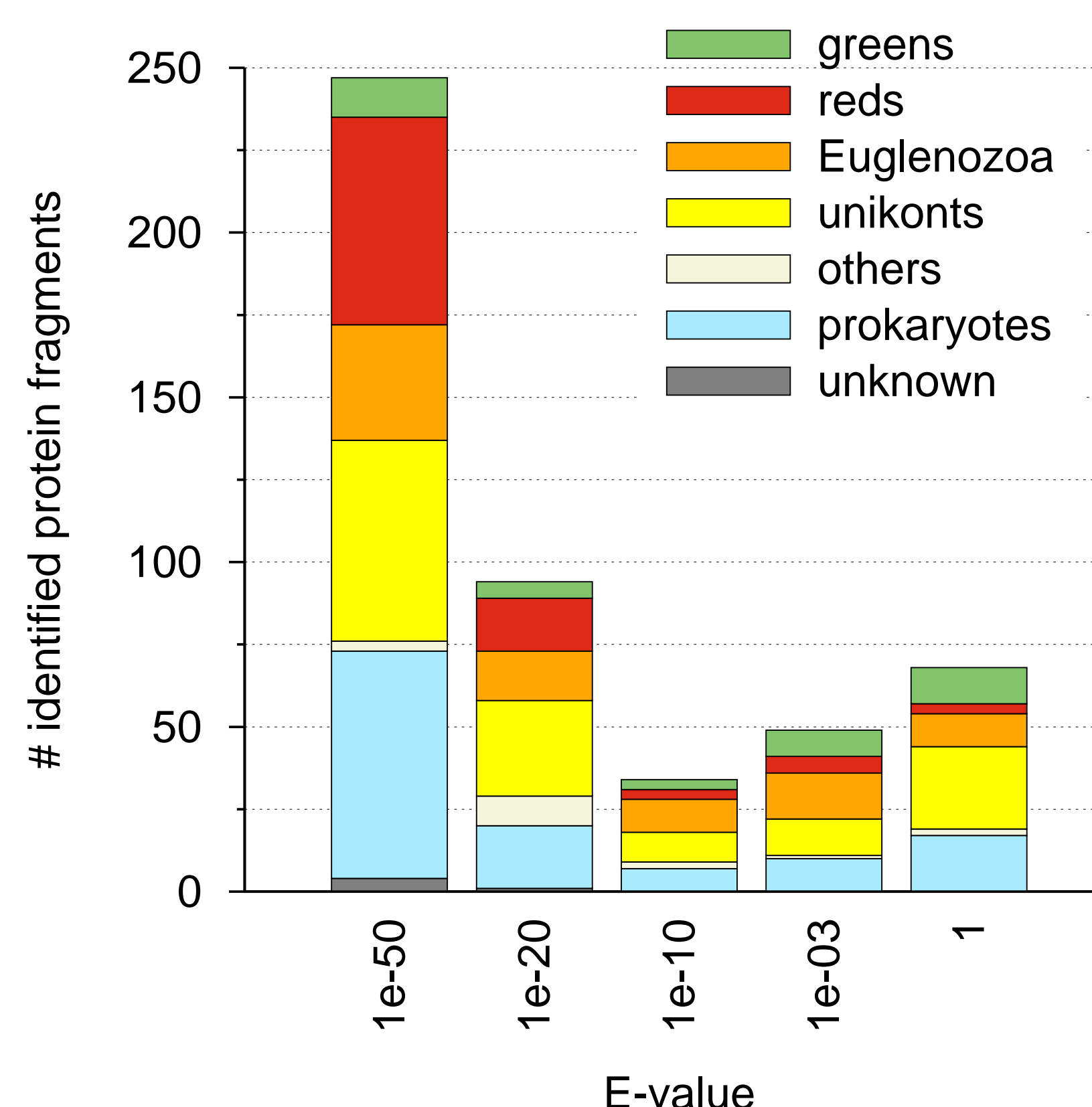


Figure 4: Taxonomic affiliation of protein fragments broken by E-value. 1e-50: $E\text{-value} \leq 1e-50$, 1e-20: $1e-50 < E\text{-value} \leq 1e-20$, 1e-10: $1e-20 < E\text{-value} \leq 1e-10$, 1e-03: $1e-10 < E\text{-value} \leq 1e-03$ and 1: $1e-03 < E\text{-value} \leq 1$.

WHEN we examine the group of organisms corresponding to the first BLAST hits, results are inconclusive. Indeed, there are relatively few organisms belonging to Euglenozoa, contrary to what is expected, and this trend persists over all ranges of E-value. Moreover, we observe a relatively large proportion of hits corresponding to red algae and allies ("reds"), while hits from green algae and land plants ("greens") are much less abundant. Though already reported in *Euglena gracilis*, this remains to be explained.

Coverage of the mitochondrial metabolic pathways

FINALLY, we searched for mitochondrial metabolic pathways present in our set of identified proteins.

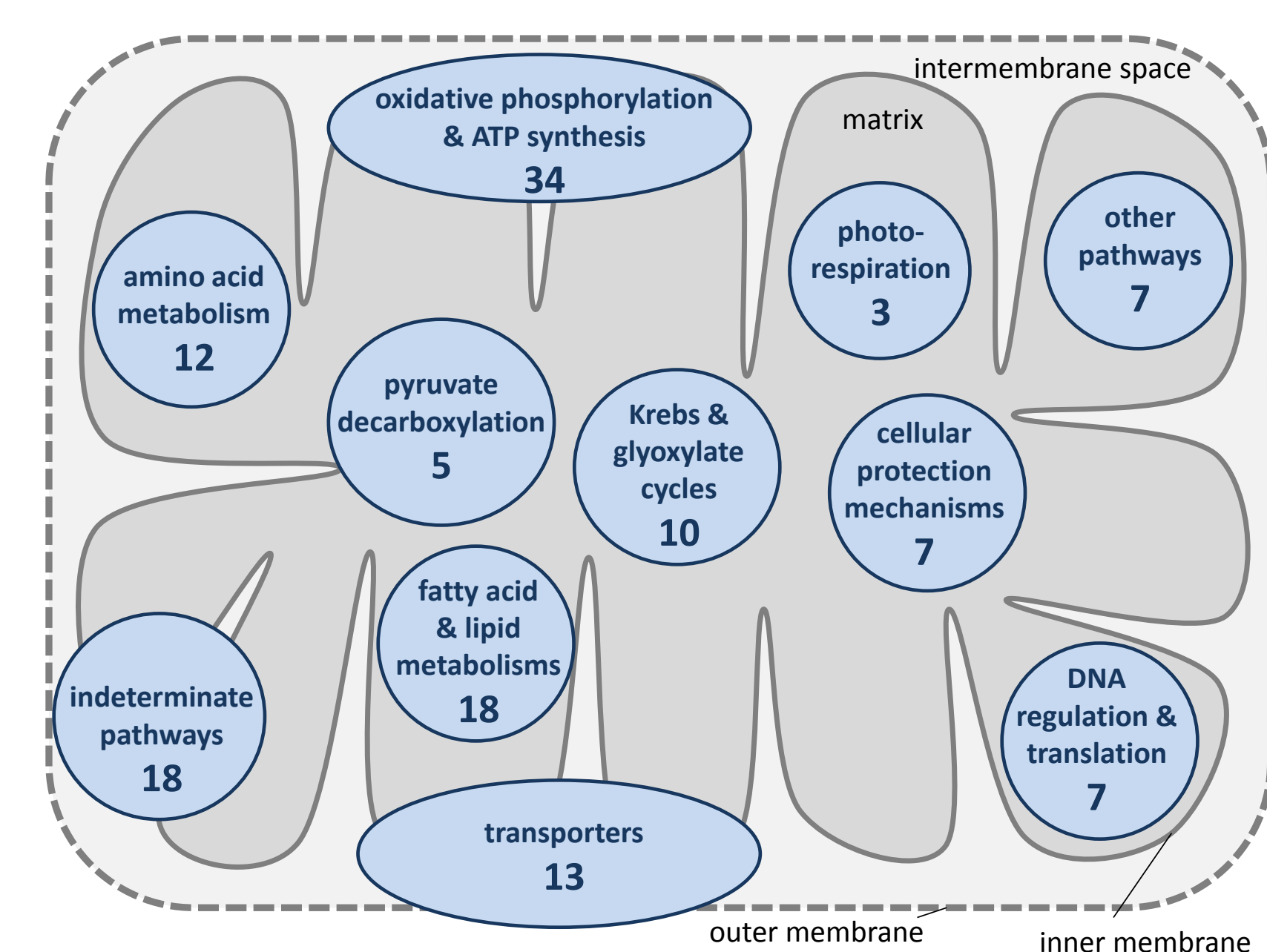


Figure 5: Coverage of the mitochondrial metabolic pathways. Number of identified proteins for each pathway.

FOR each pathway, we report the number of different identified proteins. Most mitochondrial metabolic pathways are represented by many of their proteins. The most represented pathway is oxidative phosphorylation associated with ATP synthesis (34 identified components), which was expected since this is the most important pathway in terms of components (the four complexes of the respiratory chain and ATP synthetase are mostly composed of several tens of proteins). Thus, we identify a total of 134 mitochondrial proteins involved in various metabolic pathways.

Preprocessing of the EST database

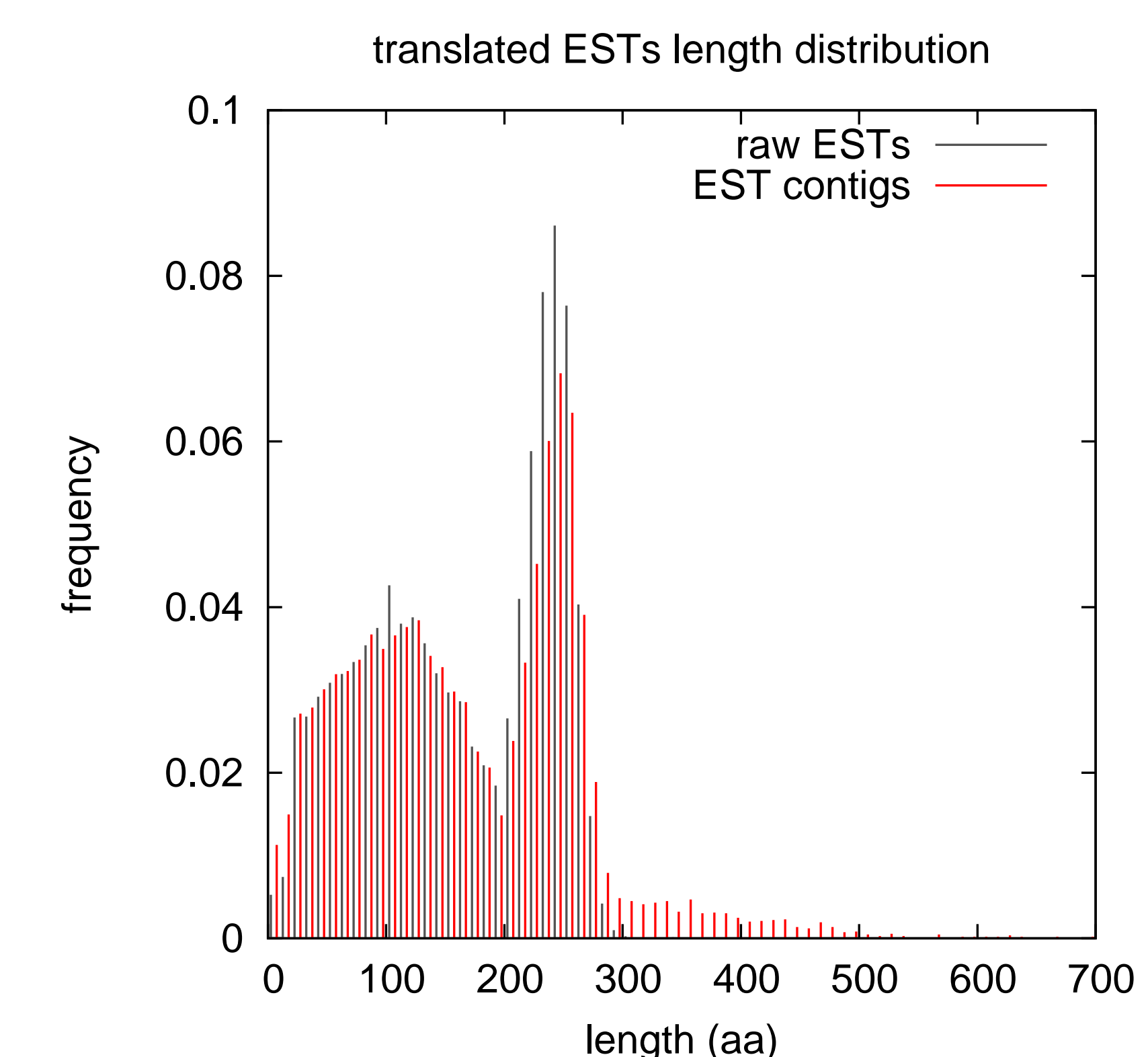


Figure 6: Translated ESTs length distribution. Lengths are those of the longest ORF for each EST.

AMONG the approaches likely to increase the number of identified proteins in the mass spectrometry analysis, the preprocessing of the public database of ESTs from *Euglena gracilis* is the easiest to implement. Indeed, assembling EST contigs with cap3 allowed us to obtain longer ESTs, thus providing better identification of protein fragments.